DOCUMENT RESUME

ED 044 063 48 FL 001 969

AUTHOR	Paquette, Andre F.; Tollinger, Suzanne
TITLE	The MLA Foreign Language Proficiency Tests for Teachers and Advanced Students: Analysis of the Performance of Native Speakers and Comparison with That of NDEA Summer Institute Participants.
INSTITUTION	Modern Language Association of America, New York, N.Y.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-6-2619
PUB DATE	Jun 68
CONTRACT	OEC-1-6-062619-1876
NOTE	44p.
EDRS PRICE	EDRS Price MF-$0.25 HC-$2.30
DESCRIPTORS	*Achievement Tests, Educational Testing, French, German, Italian, Language Instruction, Language Proficiency, Language Research, Language Skills, *Language Teachers, Language Tests, *Modern Languages, *Native Speakers, Objective Tests, Spanish, Standardized Tests, *Teacher Qualifications, Testing Programs, Verbal Tests
IDENTIFIERS	*MLA Foreign Language Proficiency Tests, NDEA Language Institutes

ABSTRACT

This report is a summary of the results of the administration of the Modern Language Association of America (MLA) Proficiency Tests to native speakers in Chile, Columbia, France, German, Italy, and Spain. The notion of a "superior" level of competence in reading, writing, aural comprehension, and speaking a second language is specified objectively through the statistical analysis of the tests administered by the MLA research associates. The tests of some 300 individual speakers of each language (French, German, Italian, and Spanish) were scored by the Educational Testing Service and compared with scores on identical tests administered to National Defense Education Act (NDEA) summer institute participants. Twenty-four tables and figures present statistical information concerning the findings. (RL)

OEC-1-6-062619-1876
TITLE VI, NDEA
SECTION 602
IIS, OE
PA 48

EDO 44063

THE MLA FOREIGN LANGUAGE PROFICIENCY TESTS FOR TEACHERS AND ADVANCED STUDENTS

ANALYSIS OF THE PERFORMANCE OF NATIVE SPEAKERS

AND

COMPARISON WITH THAT OF NDEA SUMMER INSTITUTE PARTICIPANTS

June 1968

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

by

MODERN LANGUAGE ASSOCIATION OF AMERICA

F. André Paquette, Project Director
Suzanne Tollinger, Research Assistant

FL 001 969

with the cooperation of

Educational Testing Service
Madeline Wallmark, Statistical Analysis Division

The research reported herein was performed pursuant to a contract with the Office of Education, United States Department of Health, Education, and Welfare. Contractors undertaking such projects under government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

PREFACE

This report has been prepared in partial fulfillment of Contract OEC-1-6-062619-1876, "A Continuing Survey of Foreign Language Resources of the Country through Professional Leadership in the Development and Use of Foreign Language Tests," dated 30 March 1966. The report is a summary of the results of the administration of the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students to "native speakers" in Chile, Colombia, France, Germany, and Spain.

The success of the project is due in large measure to the efforts of the MLA Research Associates, who adapted testing procedures to foreign testing conditions and planned and supervised the administration of the tests abroad:

Edward D. Allen
Ohio State University
France

Salvatore J. Castiglione
Middlebury College
Italy

Filomena del Olmo
Morris Township, New Jersey
Chile, Colombia, Spain

Gustave Mathieu
California State College
Fullerton
Germany

Robert L. Baker, Indiana University, adapted the Russian tests, but it was impossible to arrange a test administration in the Soviet Union.

The following institutions cooperated with the Research Associates in making their facilities and students available for the administration of the tests:

Escuela Central de Idiomas, Madrid, Spain

Escuela Militar, Santiago, Chile

Institut National des Sciences Appliquées (INSA), Lyons, France

Liceo-Ginnasio Dante, Liceo-Ginnasio Galileo, and Liceo Scientifico Leonardo da Vinci, Florence, Italy

Pädagogische Hochschule, Berlin, West Germany

Universidad Javeriana, Bogotà, Colombia.

MLA FOREIGN LANGUAGE PROFICIENCY TESTS
FOR
TEACHERS AND ADVANCED STUDENTS

Analysis of Performance of Native Speakers
and
Comparison with that of NDEA Summer Institute Participants

I. Introduction

While a "superior" level of competence in reading, writing, aural comprehension and speaking was defined by the Steering Committee of the Foreign Language Program of the Modern Language Association of America (MLA) as proficiency approximating that of an educated native, no formal effort had ever been made to determine the actual performance of a representative group of "native speakers" on the MLA Foreign Language Proficiency Tests for Teachers and Advanced Students (MLA Foreign Language Proficiency Tests).

Under its contract with the United States Office of Education (Contract No. OEC-1-6-062619-1876) to conduct a continuing survey of the foreign language resources of the country, and to provide professional leadership in the development and use of foreign language tests, the MLA, with the technical assistance of Educational Testing Service (ETS), undertook a large-scale investigation of the performance of groups of native speakers of French, German, Italian, and Spanish on currently active forms of the tests; a similar study of a group of Russian native speakers was planned, but, because of difficulties which arose in arranging a test administration in the Soviet Union, it was not possible to carry out the investigation as planned. It was hoped that such an investigation would provide not only data on the relative performance of native speakers, but also an insight into the strengths and weaknesses of the existing instruments, with a view to providing data which would lead to the development of improved forms of the tests.

Research associates in each language were appointed by the MLA to adapt the testing procedures to foreign testing conditions, to translate the existing test directions and simplify the recording of responses, to modify the administration of the tests as necessary, and to supervise

the administration of the tests abroad. In order to anticipate changes necessitated by conditions abroad, existing forms of the tests were experimentally administered to small groups of native speakers. In the light of this experimental "pilot" administration, materials were prepared and printed by the MLA for subsequent administration to samples of about 300 individuals in each of the four languages noted above. The research associates were also asked to develop a means for obtaining personal data on the samples tested. Each language group developed a personal data questionnaire appropriate for that language sample.

Following test administrations conducted by the research associates, the materials (Listening Comprehension, Speaking, Reading, and Writing Tests) were returned to ETS for professional scoring and transcription prior to the analysis described in this report and summarized in the set of tables and figures following the text. Table 1 presents test information that will be useful in the interpretation of the results.

II. Administration of the Tests

Although the sampling objective was to obtain for each language a representative group of educated native speakers, practical considerations imposed certain limitations which must be considered in interpreting the data. It was for this reason that each group of research associates collected personal data which, in their opinion, might be significant in the interpretation of test results for that language sample. The characteristics that were so classified are: (a) sex; (b) age, or general maturity; (c) educational level; (d) place of residence; (e) professional goal or affiliation; and (f) motivation. The results of the personal data questionnaires are summarized in Tables 2 through 5 at the end of this report.

The French Sample and Administration of the French Tests

The four skills tests in French were administered at the Institut National des Sciences Appliquées (INSA) in Lyon, France, in March, 1967. All of the people in the sample were students enrolled full-time at the institute, and most of them held the "baccalaureat".

Each examinee was offered $2.00 for participating in the study. About two-thirds of the students were from middle-class homes, the others were from a lower socio-economic level and were attending the institute on scholarships. Most of the students came from many different regions in France, but a few were from North Africa and Madagascar.

The report* on the problems and procedures of administration includes statements which add to the description of the sample:

> "Inasmuch as the purpose of INSA is to prepare engineers, the study of humanités and languages is deemphasized. Relatively little attention is given to the study of the French language and literature. Consequently, it should be kept in mind that the performance on the MLA tests would be inferior to that of a group of students from a regular French university. On the other hand, the level of language skill demonstrated by the examinees of INSA is probably typical of the 'average' native speaker of French."

Another comment in the report concerns a problem in administration and motivation:

> "In the large group of 230 students (those tested on the first day) there was laughing, talking, and jeering at the beginning of the reading and writing exam. This subsided considerably once the exam began. However, there was a considerable amount of conversation among the examinees while this particular exam was in progress. The monitors . . . seemed to have little control over the examinees."

In view of the fact that objective testing is relatively new in European countries, the above comment may indicate an expected student reaction to a testing situation which on first inspection appeared to be trivial, but after further consideration became worthy of serious attention. The results on these two tests are, in fact, consistent with those on the other two tests.

---

*Edward D. Allen, "Report on the MLA Foreign Language Proficiency Tests Validation: French"

By combining this information with the more objective information presented in Table 2, one can obtain a reasonably clear description of the French sample.

The German Sample and Administration of the German Tests

The four skills tests in German were administered at the Pädagogische Hochschule in Berlin in May, 1967. The examinees were recruited from the Pädagogische Hochschule and the Free University of Berlin and were each given the equivalent of $5.00. The recruiting procedures were carefully worked out to insure a sample that was as representative as possible in the balance of sexes and the representation of all fields of study. The report* on the details of the administration emphasizes the careful control of sampling and administration for this group. The personal data information for this sample is summarized in Table 3.

The Italian Sample and Administration of the Italian Tests

The four skills tests in Italian were administered in June, 1967, at three schools in Florence: The Liceo-Ginnasio Dante, the Liceo-Ginnasio Galileo, and the Liceo Scientifico Leonardo da Vinci. The examinees, each of whom received approximately $2.50, were students completing the tenth, eleventh, and twelfth years of school.

In this administration the technical problems encountered with the equipment required for the Listening Comprehension and the Speaking Tests had a noticeable effect upon the test results. The first two schools mentioned above tested 50 students each and managed the Listening Comprehension Test administration with one tape recorder each. The third school, which tested about 200 students, lacked the kind and quantity of equipment desired for testing this large a number at one time, and had to resort to using the school's public address system for the Listening Comprehension Test. There is some indication that this arrangement was not by any means ideal from the administration point of view, but was good enough to serve the purposes of this study.

---

*G. Mathieu, "Report on the MLA Foreign Language Proficiency Tests Validation Study: German"

Since only ten tape recorders could be borrowed for the Speaking Test, a "makeshift substitute for a language laboratory"* was set up each afternoon at the Liceo Scientifico Leonardo da Vinci, and the examinees were tested after the regular school session in relays of ten students at a time. With this arrangement, 279 students took the Speaking Test, but of the 279 tapes only 227 were scorable. The scorers reported some problems with background noises and equipment problems, but since the deletion of unscorable tapes was probably random with respect to the students' speaking abilities, the results of the 227 scorable tapes are usable in this study.

The summary of the personal data information, presented in Table 4, indicates that this sample is less mature than the samples for French, German, and Spanish.

The Spanish Samples and the Administration of the Spanish Tests

The skills tests in Spanish were administered in three countries: Chile, Colombia, and Spain. The summary of the personal data for the total sample is presented in Table 5, but since the three subgroups are analyzed separately as well as together, additional information is needed to characterize each subgroup.

Since the largest and best language laboratory facilities in Chile were located in the Escuela Militar in Santiago, this institution was used as a testing center and the source of examinees for the Chilean sample. At the request of the director of the school, who also served as foreign coordinator, the compensation of $790 was paid to the school. The students received no compensation for their participation. The installation consisted of two laboratories with a total of thirty booths, each equipped with a tape recorder and earphones and controlled from a central console. The tests were administered in February, 1967, to one hundred military cadets, most of whom were between the ages of 18 and

---

*Salvatore J. Castiglione, "Report on the MLA Foreign Language Proficiency Tests Validation Study: Italian"

22 years. The report* on the administration details indicates that various problems were encountered that probably had an effect on the test results for this group. In each of the testing rooms the proctor was the senior cadet, who was also one of the examinees. The report states that "one group of twenty-five was completely incorrigible to the extent that their captain exercised little control over them" and the "Speaking tapes recorded by this group will be invalid because of the cadets' lack of seriousness and the fact that they thought nothing of making comments to each other, giggling and whistling during the course of the Speaking Test" and these tapes were not scored.

In Colombia the administration of the Spanish tests was held in February, 1967, in the Universidad Javeriana, a Catholic University in Bogotá. At the request of the vice consul and officer in charge of the financial affairs at the United States Embassy, the compensation of $500 was given to the university. The language laboratory facilities here were excellent, consisting of one hundred positions, twenty per room, each visible from the control unit. The director of the foreign language department of the university had considerable difficulty in recruiting one hundred students who were willing to give up their free time to participate in the project, but finally succeeded in assembling a Colombian sample including 68 seminarians from the Chapinero Seminary and 31 male and female psychology students from the Universidad Javeriana. A minor incident (the floor plug that controlled the current in ten booths was accidentally knocked out) resulted in the failure of ten examinees to answer three items on the Speaking Test, but since this amount of error is less than 0.03 in the sample mean, it may be overlooked.

In Spain the tests were administered in June, 1967, at the Escuela Central de Idiomas. The laboratory facilities here consisted of 132 booths in five rooms. Each room had a master console and the equipment permitted both listening and speaking activities. The director of the Escuela Central requested that the compensation of $440 for the school be paid in American dollars and approved the payment of Kennedy

---

*Filomena del Olmo, "Report on the MLA Foreign Language Proficiency Tests Validation Study: Spanish"

half-dollars to the students. The sample consisted of 109 students from the Instituto Internacional and the Escuela Central de Idiomas. As was planned, there were few male participants in this group (only 11) since the majority of the participants in Chile and Colombia had been males. The age range for this group was from 15 to 58. At the end of the testing the participants were asked several questions which may be considered in the interpretation of the test results:

1. Have you had previous experience in a language laboratory? Yes: 17 No: 91
2. Did the dialogues and newscast seem too long to you? Yes: 42 No: 66
3. Have you ever taken this type of multiple-choice exam before? Yes: 33 No: 74
4. Was the Reading Test difficult or easy? Very difficult: 2 Difficult: 48 Easy: 56

In the Spanish report, Filomena del Olmo points out in detail the limitations of the data collected from the Spanish samples. Because two of these comments are particularly relevant in the interpretation of the results for these groups, they are repeated here:

> "In this country (the United States) the examinees . . . are motivated to do their best since they know that the tests will be scored and they will be officially recorded. Taking the test is a significant professional activity . . . The examinees in foreign countries, naturally, do not approach the testing with the same motivation . . ."
>
> "With regard to the Speaking Test, I feel that the Spanish-speaking examinees in this country (the United States), insofar as their linguistic sophistication allows, are 'on guard' and using a level of language that Martin Joos . . . describes as 'good, standard, mature, consultative' . . . The native speakers abroad used a 'casual, provincial, fair' level of language and still others used an 'intimate, popular' level . . ."

III. Processing of the Data

In order to facilitate the statistical work required for the analysis of the data, a staff of clerical workers at ETS transferred the responses for the Listening and Reading Tests from the improvised answer sheets and test booklets to standard machine-scorable answer sheets.

The Writing and Speaking Tests, both of which require scoring by language specialists, were scored along with the 1967 NDEA* Summer Institute answer sheets and Speaking tapes by the staff of professional scorers employed by ETS for this part of the MLA FL Proficiency Tests processing. Some of the comments that the professional scorers made on the problems they encountered in their work are a repetition of those made by the administrators and noted in Section II above. The mimicry sections of the Speaking Tests requires the student to repeat phrases exactly as heard from the master tape. The professional scorers are required to rate as "incorrect" a sound which is not a reproduction of the master voice, even though the pronunciation may be correct. Additional ones involved poor identification of Speaking tapes, confusing background noises on the Speaking tapes, and specific items which seemed inappropriate for native speakers.

When all of the data had been recorded in machine-readable form, the answer sheets were scored and the scores collated with the coded personal data information. The collated information was then processed to produce rosters, distributions, item analyses, and intercorrelation tables.

IV. Analysis and Interpretation

All results are presented in terms of converted scores. In order to avoid invalid comparisons it is necessary to point out certain characteristics of the scales used for reporting scores on the MLA Proficiency Tests. The converted score scale for each test for each

---

*Foreign language summer institutes operated under the auspices of the National Defense Education Act (NDEA).

language was established on the first form of that test by merely adding 20 points to the "raw" score. This was done so that when new forms were introduced and equated to the original form it would be unlikely that the resulting converted score range for those forms would have a minimum reportable score less than 0. This means that no comparisons of converted scores across languages or among tests within a language are valid. Comparisons are limited to those among groups taking the same test in the same language.

Distributions of Converted Scores - The performance of the native speaker groups is summarized in the form of distributions of converted scores in Tables 6-9. Each table includes summary statistics (number of cases, mean, and standard deviation) for the native speaker group and, for purposes of comparison, the corresponding statistics for the 1961-65 NDEA Summer Institute groups.* The latter information was taken from the score interpretation leaflet that accompanies the score reports for the MLA Proficiency Testing Program. A comparison of the statistics for each native speaker group on each test with those for the corresponding NDEA group shows that in every case the native speaker group has a higher level of performance (higher mean) and that the group is less variable (smaller standard deviation).

Table 10 presents the summary statistics for the three Spanish subgroups. As one might expect from the available descriptive information on the samples, the sample from Chile has a somewhat lower level of performance than the samples from Colombia and Spain, but even the Chilean group has a significantly higher level of performance than that of the 1961-65 posttest group.

To simplify the interpretation of the distributions, the information has been condensed in Tables 11-14 to show converted scores for

---

*Foreign language summer institutes operated under the control of the National Defense Education Act (NDEA) for the purpose of providing an intensive period of instruction for teachers and advanced students in foreign languages. Most institutes test their participants at the beginning and at the end of the summer training period.

selected percentile ranks for the native speaker groups and also for the 1961-65 and 1966 NDEA pretest and posttest groups. Table 15 presents similar information for the three Spanish subgroups. These tables show that the 10th percentile for the native speaker groups is consistently higher than the 50th percentile for the NDEA posttest groups. These comparisons become more obvious when one examines the corresponding graphs presented in Figures 1-4. Each figure presents a graphical summary for one of the four skills tests, showing for each language the relative performance of the 1961-65 NDEA pretest and posttest groups, the 1966 NDEA pretest and posttest groups, and the native speaker groups. The three Spanish subgroups are also included in the graph. Each bar shows the 10th, 25th, 50th, 75th, and 90th percentiles for one group. For example, in Figure 1 one can see that the 10th percentile of the Italian native speaker sample on the Italian Listening Comprehension Test is on the same level as the 90th percentile of the 1966 NDEA posttest group. A word of caution is in order here. Comparisons across languages for a given skills test or within a language across the four skills tests are not valid, except in a very general sense. It is apparent that all native speaker groups have a significantly higher level of performance on all of the skills tests. In Figure 2, the graph for the Speaking Tests, the bar for the native speaker group in German does not overlap those for the NDEA groups, whereas those for the native speaker groups in French, Italian, and Spanish show a slight overlap, indicating that for these groups the lower-scoring native speakers are on the same level of performance as the higher-scoring NDEA posttest examinees. This is not surprising when one notes that the German native speaker sample consisted of mature students at a university in Berlin and that this group was more highly motivated than the other native speaker groups.

Intercorrelation Tables - Table 16 presents the intercorrelation tables for the four native speaker groups. Since these tables must be based on matched cases (those having all four scores), the sample N's are lower than those for the distributions. The intercorrelation table for the original French test analysis sample based on 1961 NDEA posttest results is also given for general comparison. The latter table shows rather high correlations ranging from .736 (Speaking vs. Reading) to .858 (Reading vs. Writing). Note the very low values for the French native

speaker sample, particularly for the correlations between Speaking and the other tests. These low values can be attributed to the very small standard deviations. The same characteristics are apparent in the tables for German, Italian, and Spanish.

VI. Item Analysis

Item analysis was performed on the MLA Foreign Language Proficiency Tests for the native speaker samples in order to provide detailed information about the performance of each test item for the special native speaker group and to show whether or not certain items require revision. This kind of analysis is more technical than the information presented above, and, therefore, a brief explanation and some definitions are appropriate as an introduction.

Item analysis provides detailed statistical information describing how a particular item functioned in a particular test for a particular sample. It provides information about the difficulty of the item for that sample, the relative attractiveness of the options, and the correlation of the item score with the score on the total test, or the criterion score. In the method of analysis used in this investigation the criterion score is converted in such a way that the distribution is normal with a mean of 13.0 and a standard deviation of 4.0. Then the following statistics are computed.

P+: the proportion of the sample answering the item correctly.

Δ, or delta: The index of difficulty of the item, based upon the P+. Deltas range from less than 6.0 (if more than 95% answer the item correctly, the delta is not computed and can only be estimated as less than 6) to more than 20.0 (if less than 5% answer correctly, the delta is not computed and can only be estimated as greater than 20). A delta of 13.0 means that 50% of the sample answered the item correctly.

r-biserial: an index of discrimination measuring the extent to which examinees who scored high on the criterion score tend to answer the item correctly, and those who score low tend to answer incorrectly. For foreign language tests an item analysis based on a sample for whom the test was designed will typically yield

r-biserials that are high compared with those for tests on other subject matter. Values above .70 are not infrequent for the language tests, but are practically non-existent for tests in other fields.

Table 17 is included for the record, for it shows the specific identification printed on the detailed item analysis and will permit test specialists to complete a detailed evaluation of the tests item by item. Tables 18-22 present the frequency distributions of the deltas and r-biserials for the native speaker groups, identified by NS, and for the original test analysis samples, identified by TA. The test analysis samples were selected from the NDEA pretest and posttest groups. The distributions for the test analysis samples are typical for tests appropriate for the group: the mean delta is approximately 12-13 and the mean r-biserial above .50. The distributions for the native speaker groups show that these tests are too easy to be used as reliable measuring instruments for similar samples. The consistently lower values for the mean r-biserial result from the greater homogeneity of the native speaker samples.

Since "items" 41-53 (the rating scales) on the Speaking Tests cannot be analyzed in the same way as the other items, the results of the ratings are summarized in Tables 23 and 24. Each table is arranged to show the ratings by picture (as indicated by I, II, III) and criterion for rating (Vocabulary, Pronunciation, Structure, and Fluency). Mean ratings are given for each classification. It is interesting to note that the sample from Spain has significantly high ratings on pronunciation, but the means on the other ratings are between those of the Chilean and Colombian samples. There is a suggestion here that the professional scorers have a slight bias in favor of the pronunciation characteristic of the sample from Spain. Some allowance must be made for the attitude of the South American samples toward this test, and the scorer bias may not be as significant as it appears to be.

VII. Conclusions

In spite of all the problems encountered in obtaining samples that could be classified as "educated native speakers" and in administering the tests in countries where language laboratory facilities and objective testing are relative novelties, the results of this investigation can serve as a guide in defining a "superior" level of competence as measured by the skills tests of the MLA Proficiency Tests for Teachers and Advanced Students. The comments of the professional staff who worked on the sampling, administration, and professional scoring will be valuable in reviewing the effectiveness of the existing instruments and providing a basis for revision. In interpreting the results on the Speaking Tests, one must bear in mind the special problems of administration and scoring involved. The equipment problems of the Italian and Spanish samples were greater than those experienced in a routine national administration of the tests in this country. The scoring of the mimicry sections for the native speakers suggests a weakness in the setting of standards for scoring this type of item. The native speaker's approach to the Speaking Test items is different from that of the American student of a foreign language. But all of these factors do not change the basic conclusion of this study that the native speaker groups performed at a level considerably higher than that of the NDEA posttest groups, but that there is some overlap in performance, suggesting that the best among the NDEA participants approach the "educated native speaker" in competence.

All of the input data has been retained in the form of rosters showing identification number, coded personal data information, and converted scores on the four skills tests. Examinees with incomplete information are identified by asterisks. The detailed item analysis, which has been turned over to the Modern Language Association, can be used for an item-by-item analysis of each test. Both of these records include information for additional research.

List of Tables and Figures

Table 1	Test Information
Tables 2-5	Description of Samples: Summary of Personal Data Information
Tables 6-9	Distributions of Converted Scores (French, German, Italian, Spanish)
Table 10	Summary Statistics for the Spanish Subgroups
Tables 11-15	Converted Scores Corresponding to Selected Percentile Ranks for Native Speakers and for NDEA Groups (French, German, Italian, Spanish, and Spanish Subgroups)
Figures 1-4	Comparison of the Performance of Native Speakers and NDEA Groups (Listening Comprehension, Speaking, Reading, Writing)
Table 16	Intercorrelation Tables (French - Native Speaker Sample, French - Original Test Analysis Sample, German, Italian, Spanish)
Table 17	Item Analysis Specifications and Identification
Tables 18-22	Frequency Distributions of Item Statistics (French, German, Italian, Spanish, Spanish Subgroups)
Table 23	Summary of Ratings on the Speaking Tests (French, German, Italian)
Table 24	Comparison of the Ratings for the Three Spanish Subgroups on the Spanish Speaking Test

Table 1. Test Information

Modern Language Association Foreign Language Proficiency Tests
for Teachers and Advanced Students

Test Title	Time Limit	Scorable Units	Maximum Raw Score	Converted Score Range
LISTENING COMPREHENSION				
French	Approx. 20 min.	36	36	20 - 58
German	"	36	36	6 - 58
Italian	"	36	36	20 - 56
Spanish	"	36	36	22 - 63
Speaking: Part A		(20)	(20)	
Part B-1		(20)	(20)	
Part B-2		(1)	(5)	
Part C		(12)	(60)	
French	Approx. 15 min.	53	105	17 - 128
German	"	53	105	0 - 141
Italian	"	53	105	20 - 125
Spanish	"	53	105	17 - 116
Reading				
French	40 min.	50	50	19 - 69
German	"	50	50	16 - 70
Italian	"	50	50	20 - 70
Spanish	"	50	50	20 - 69
Writing				
French	45 min.	60	60	18 - 77
German	"	60	60	24 - 83
Italian	"	60	60	20 - 80
Spanish	"	60	60	25 - 91

The Listening Comprehension Tests and the Speaking Tests are self-timing.

The Writing Tests and the Speaking Tests are scored by professional scorers. Parts A and B-1 of the Speaking Tests are rated as right or wrong, Parts B-2 and C are rated on a five-point scale - (1-5).

The converted score scales were established independently for each test within each language. There was no attempt to establish scales that would permit direct comparisons of converted scores for any pair of tests.

Table 2. Description of the FRENCH Sample

(Summary of Personal Data Information)

Sample Characteristic	Number of Cases	Per Cent of Sample
Number of Observations	307	
Sex		
(A) Male	250	81.43
(B) Female	19	6.19
(-) No response	38	12.38
Age		
(A) 17-21 Years	255	83.06
(B) 22-25 Years	23	7.49
(-) No response	29	9.45
Educational Level		
(A) First Year (INSA)	156	50.81
(B) Second Year (INSA)	68	22.15
(C) Third Year (INSA)	24	7.82
(D) Fourth Year (INSA)	20	6.51
(-) No response	39	12.71
Residence (0-5 Years of Age)		
(A) South France	104	33.88
(B) North France	136	44.30
(C) Paris	29	9.45
(D) Foreign, French-speaking	9	2.93
(-) No response	29	9.45
Residence (from 5 to 15 years of age)		
(A) South France	104	33.88
(B) North France	137	44.63
(C) Paris	33	10.75
(D) Foreign, French-speaking	4	1.30
(-) No response	29	9.45
Residence (after 15 years of age)		
(A) South France	104	33.88
(B) North France	132	43.00
(C) Paris	39	12.70
(D) Foreign, French-speaking	1	.33
(-) No response	31	10.09

NOTE: The letters in parentheses are the identification codes appearing on the rosters. The same comment applies to Tables 3, 4, and 5.

Table 3. Description of the GERMAN Sample

(Summary of Personal Data Information)

Sample Characteristic	Number of Cases	Per Cent of Sample
Number of Observations	311	
Sex		
(A) Male	142	45.66
(B) Female	168	54.02
(-) No response	1	.32
Age		
(A) 18-21 Years	172	55.30
(B) 22-25 Years	114	36.66
(C) 26 Years or Older	24	7.72
(-) No response	1	.32
Educational Level		
(A) 1-4 semesters	214	68.81
(B) 5-10 semesters	75	24.12
(C) 11 or more semesters	21	6.75
(-) No response	1	.32
Residence		
(A) Berlin	172	55.30
(B) North Germany	76	24.44
(C) South Germany	13	4.18
(D) East Germany	17	5.47
(E) Other	32	10.29
(-) No response	1	.32
Major Field		
(A) Natural and Physical Sciences	46	14.79
(B) Humanities, Fine Arts	171	54.99
(C) Law, Political Science, Economics, Mathematics, Sociology, Psychology	76	24.44
(D) Home Economics, Physical Education	16	5.14
(-) No response	2	.64

Table 4. Description of the ITALIAN Sample

(Summary of Personal Data Information)

Sample Characteristic	Number of Cases	Per Cent of Sample
Number of Observations	286	
Age		
(A) 16-17 Years	153	53.50
(B) 18-19 Years	128	44.75
(C) 20 Years	3	1.05
(-) No response	2	.70
Educational Level		
(A) Liceo Classico	106	37.06
(B) Liceo Scientifico	178	62.24
(-) No response	2	.70
Residence after 5 Years of Age		
(A) Outside of Tuscany	31	10.84
(B) In Tuscany	253	88.46
(-) No response	2	.70

Table 5. Description of the SPANISH Sample

(Summary of Personal Data Information)

Sample Characteristic	Number of Cases	Per Cent of Sample
Number of Observations	308	
Age		
(A) 15-18 Years	62	20.13
(B) 19-22 Years	165	53.57
(C) Over 22 Years	78	25.33
(-) No response	3	.97
Educational Level		
(A) 2-3 Years Bachillerato	7	2.27
(B) 4-5 Years Bachillerato	46	14.94
(C) 6-7 Years Bachillerato	127	41.23
(D) Attending or Completed University	114	37.01
(E) Other	11	3.57
(-) No response	3	.97
In or Out of School		
(A) In School	216	70.13
(B) Not in School	69	22.40
(-) Not Indicated	23	7.47
Professional Standing		
(A) Military Student	99	32.14
(B) Seminary Student	55	17.86
(C) Professional	13	4.22
(D) Semi-professional	76	24.68
(E) Other	3	.97
(-) No response	62	20.13
Residence		
(A) Spain (age range: 15 to 58 years) 11 Male Students 98 Female Students	109	35.39
(B) Chile (military cadets)	100	32.47
(C) Colombia 68 Jesuits, Chapinero Seminary 31 Male and Female Students, Universidad Javeriana	99	32.14

Table 6

Frequency Distributions of Converted Scores: FRENCH

LISTENING COMPREHENSION		SPEAKING		READING		WRITING	
Score (20-58)	Percent Below	Score (17-128)	Percent Below	Score (19-69)	Percent Below	Score (18-77)	Percent Below
58	90.2	120	98.3	68	97.7	76	99.7
57	72.3	117	93.8	67	92.5	75	98.7
56	52.3	114	90.0	66	84.6	74	94.8
54	33.3	111	85.5	65	71.1	73	86.9
53	21.1	108	71.3	64	54.4	72	77.7
52	13.0	105	54.3	63	43.0	71	66.6
51	7.7	102	38.1	62	30.2	70	56.7
50	5.6	99	27.0	61	22.0	69	48.2
49	3.2	96	18.3	60	14.8	68	38.4
48	1.4	93	13.8	59	11.5	67	29.2
47	1.1	90	8.3	58	7.9	66	24.9
46	1.1	87	4.5	57	5.6	65	20.3
45	0.4	84	2.4	56	4.6	64	14.8
44	0.4	81	0.3	55	2.0	63	9.8
43	0.4	78		54	1.0	62	6.6
42	0.4			53	0.3	61	4.6
41				52	0.3	60	3.0
				51		59	2.6
						58	2.0
						57	1.6
						56	1.0
						55	0.3
						54	0.3
						53	
Number of Cases	285		289		305		305
Mean	54.5		102.8		62.6		68.1
Standard Deviation	2.5		8.5		3.2		4.1
CORRESPONDING STATISTICS FOR 1961-65 NDEA							
Pretest N	7698		7413		7699		7699
Mean	38.1		68.2		43.0		42.4
Standard Deviation	8.7		18.0		10.5		12.5
Posttest N	7853		7678		7852		7908
Mean	42.8		80.2		45.3		45.2
Standard Deviation	8.4		16.1		10.3		12.3

Table 7

Frequency Distributions of Converted Scores: GERMAN

LISTENING COMPREHENSION		SPEAKING		READING		WRITING	
Score (6-58)	Percent Below	Score (0-141)	Percent Below	Score (16-70)	Percent Below	Score (24-83)	Percent Below
58	91.3	139	97.3	70	97.1	82	99.0
57	72.7	138	93.3	69	91.0	81	97.7
56	49.3	137	88.0	68	75.8	80	95.2
55	49.3	136	88.0	67	58.4	79	88.7
54	30.7	135	76.9	66	45.8	78	81.3
53	17.3	134	67.9	65	31.9	77	72.3
52	17.3	133	54.5	64	20.3	76	59.7
51	10.3	132	54.5	63	20.3	75	50.0
50	5.0	131	44.1	62	13.2	74	38.1
49	5.0	130	37.1	61	10.0	73	28.4
48	1.7	129	28.1	60	6.5	72	21.3
47	0.7	128	28.1	59	4.2	71	15.5
46	0.7	127	24.4	58	2.9	70	12.3
45	0.7	126	17.4	57	1.6	69	9.0
44		125	14.0	56	1.0	68	6.5
		124	14.0	55	0.3	67	4.2
		123	8.0			66	1.9
		122	5.0	50		65	1.3
		121	5.0			64	0.6
		120	2.7				
		119	2.3			59	0.3
		118	1.3				
		117	1.3			54	
		116	1.3				
		115	0.7				
		114	0.3				
		103					
Number of Cases	300		299		310		310
Mean	54.4		130.4		65.0		74.1
Standard Deviation	2.8		5.4		3.1		3.9
CORRESPONDING STATISTICS FOR 1961-65 NDEA							
Pretest N	1757		1733		1759		1759
Mean	39.3		80.2		45.6		47.1
Standard Deviation	9.2		17.4		11.8		16.3
Posttest N	2086		2053		2086		2087
Mean	43.2		87.8		49.1		50.1
Standard Deviation	9.1		18.9		10.9		14.8

Table 8

Frequency Distributions of Converted Scores: ITALIAN

LISTENING COMPREHENSION		SPEAKING		READING		WRITING	
Score (20-56)	Percent Below	Score (20-125)	Percent Below	Score (20-70)	Percent Below	Score (20-80)	Percent Below
55	98.7	125	97.4	69	98.2	77	99.3
54	96.5	124	96.9	68	92.6	76	96.5
53	86.1	123	91.2	67	84.9	75	92.3
52	68.7	122	85.9	66	70.5	74	85.6
51	52.6	121	78.4	65	59.6	73	77.5
50	31.7	120	68.7	64	47.4	72	65.3
49	19.6	119	61.2	63	36.8	71	54.7
48	10.9	118	48.0	62	27.7	70	41.1
47	4.8	117	41.4	61	19.3	69	32.3
46	2.2	116	34.8	60	13.0	68	26.0
45	0.9	115	26.0	59	7.4	67	17.5
44	0.9	114	20.3	58	5.3	66	10.2
43	0.4	113	17.2	57	4.2	65	7.7
42		112	14.1	56	3.5	64	4.9
		111	10.6	55	2.1	63	3.2
		110	5.3	54	1.1	62	2.5
		109	4.4	53	0.7	61	0.7
		108	3.5	52	0.7	60	0.4
		107	3.5	51	0.4		
		106	2.6			54	
		105	2.6	46			
		104	2.2				
		103	1.3				
		102	1.3				
		101	1.3				
		100	1.3				
		99	0.9				
		98	0.9				
		91	0.4				
		90					
Number of Cases	230		227		285		285
Mean	50.3		116.7		63.2		69.8
Standard Deviation	2.2		5.1		3.4		3.6
CORRESPONDING STATISTICS FOR 1961-65 NDEA							
Pretest N	114		112		114		114
Mean	40.5		89.7		45.3		52.3
Standard Deviation	6.2		15.3		11.1		14.1
Posttest N	117		115		117		114
Mean	40.7		98.7		48.4		56.0
Standard Deviation	6.2		13.2		10.8		14.1

Table 9

Frequency Distributions of Converted Scores: SPANISH (Total Group)

LISTENING COMPREHENSION		SPEAKING		READING		WRITING	
Score (22-63)	Percent Below	Score (17-116)	Percent Below	Score (20-69)	Percent Below	Score (25-91)	Percent Below
63	98.9	114	99.6	68	98.4	90	98.7
62	98.2	112	99.3	67	94.8	88	96.4
61	97.1	110	96.3	66	89.5	86	90.5
60	94.2	108	93.4	65	79.3	84	86.9
59	84.0	106	90.1	64	72.8	82	74.8
58	84.0	104	83.4	63	68.2	80	62.3
57	74.2	102	76.1	62	59.7	78	46.9
56	65.1	100	68.0	61	51.5	76	35.4
55	53.5	98	61.0	60	44.6	74	30.2
54	44.4	96	53.3	59	37.0	72	21.0
53	37.8	94	45.2	58	32.8	70	12.1
52	27.6	92	33.8	57	28.9	68	7.9
51	23.6	90	26.5	56	23.3	66	4.6
50	23.6	88	22.8	55	16.1	64	3.3
49	16.7	86	17.6	54	12.8	62	1.3
48	12.0	84	12.9	53	9.5	60	0.7
47	6.9	82	7.7	52	8.2		
46	4.7	80	5.9	51	6.9		
45	3.6	78	4.4	50	6.2		
44	2.5	76	4.0	49	4.9		
43	2.5	74	2.9	48	3.3		
42	2.5	72	2.6	47	3.0		
41	1.8	70	1.8	46	2.6		
40	1.5	68	1.5	45	0.7		
39	1.1	66	0.7	44	0.3		
38	0.7			43			
		60					
31							
Number of Cases	275		272		305		305
Mean	53.3		81.8		59.4		77.1
Standard Deviation	4.9		10.0		5.2		6.4
CORRESPONDING STATISTICS FOR 1961-65 NDEA							
Pretest N	7390		7201		7381		7378
Mean	39.8		68.1		42.2		46.5
Standard Deviation	8.0		18.7		10.1		13.8
Posttest N	7418		7287		7506		7508
Mean	42.4		78.7		44.7		51.0
Standard Deviation	7.6		16.0		9.4		13.1

Table 10. Summary Statistics for the SPANISH Subgroups

Test	CHILE	COLOMBIA	SPAIN
LISTENING COMPREHENSION			
Number of Cases	80	90	100
Mean Converted Score	51.7	55.0	53.1
Standard Deviation	3.9	3.9	5.7
SPEAKING			
Number of Cases	79*	90	103
Mean Converted Score	86.3	96.2	98.5
Standard Deviation	8.5	8.1	7.3
READING			
Number of Cases	100	95	105
Mean Converted Score	58.3	60.2	59.5
Standard Deviation	4.6	5.1	5.7
WRITING			
Number of Cases	95	95	105
Mean Converted Score	75.2	78.9	76.7
Standard Deviation	5.8	6.0	6.7

*These statistics are based on the cases used for the item analysis. The type of item analysis used in this study requires sample N's that are multiples of five. In the Chilean sample for Speaking all available scores were used because the sample was so small. Even though this sample N is not a multiple of five, the item statistics and sample statistics are not affected.

Table 11

Converted Scores Corresponding to Selected Percentile Ranks: FRENCH

LISTENING COMPREHENSION	1961-65 NDEA Pretest	1961-65 NDEA Posttest	1966 NDEA Pretest	1966 NDEA Posttest	Native Speakers
Percentile Ranks	Converted Scores				
90	52	54	52	54	57
75	46	51	48	51	57
50	37	43	40	46	55
25	32	36	34	40	53
10	29	32	30	33	51
SPEAKING					
90	93	102	83	90	114
75	80	91	76	83	108
50	67	81	69	76	104
25	55	70	59	69	98
10	46	60	51	63	90
READING					
90	59	61	58	61	66
75	51	54	51	54	65
50	42	45	43	46	63
25	35	38	35	40	61
10	31	33	30	35	58
WRITING					
90	61	62	58	63	73
75	52	55	51	56	71
50	42	46	42	48	69
25	33	36	33	40	66
10	27	29	24	32	63

Table 12

Converted Scores Corresponding to Selected Percentile Ranks: GERMAN

LISTENING COMPREHENSION	1961-65 NDEA Pretest	1961-65 NDEA Posttest	1966 NDEA Pretest	1966 NDEA Posttest	Native Speakers
Percentile Ranks	Converted Scores				
90	53	56	55	56	57
75	48	51	51	53	57
50	39	45	43	47	56
25	32	36	34	40	53
10	29	31	28	34	50
SPEAKING					
90	104	113	121	120	137
75	92	100	106	108	134
50	81	88	93	95	132
25	70	76	81	86	127
10	58	65	66	77	123
READING					
90	64	65	66	66	68
75	55	59	61	61	67
50	45	49	51	54	66
25	36	40	41	45	64
10	32	35	35	39	61
WRITING					
90	71	71	67	75	79
75	61	63	60	69	77
50	47	51	51	59	75
25	33	40	40	48	72
10	26	30	29	40	69

Table 13

Converted Scores Corresponding to Selected Percentile Ranks: ITALIAN

LISTENING COMPREHENSION	1961-65 NDEA Pretest	1961-65 NDEA Posttest	1966 NDEA Pretest	1966 NDEA Posttest	Native Speakers
Percentile Ranks	Converted Scores				
90	50	49	47	47	53
75	45	47	45	44	52
50	41	42	39	40	50
25	36	36	35	35	49
10	32	32	33	30	47
SPEAKING					
90	110	115	106	113	122
75	100	110	100	106	120
50	92	101	91	100	118
25	82	90	75	92	114
10	67	82	66	87	110
READING					
90	61	64	57	60	67
75	56	59	51	56	66
50	46	48	42	46	64
25	36	41	33	40	61
10	32	34	31	35	59
WRITING					
90	71	74	68	69	74
75	67	69	62	62	72
50	53	58	52	56	70
25	42	46	42	44	67
10	33	37	35	37	65

Table 14

Converted Scores Corresponding to Selected Percentile Ranks: SPANISH

LISTENING COMPREHENSION	1961-65 NDEA Pretest	1961-65 NDEA Posttest	1966 NDEA Pretest	1966 NDEA Posttest	Native Speakers
Percentile Ranks	Converted Scores				
90	52	53	52	55	59
75	47	49	47	51	57
50	39	43	40	45	54
25	34	37	34	40	50
10	30	32	31	35	47
SPEAKING					
90	94	100	75	93	105
75	81	90	66	84	101
50	67	79	56	76	95
25	55	69	46	70	89
10	45	59	38	61	83
READING					
90	58	59	57	59	66
75	50	52	49	54	64
50	42	44	41	46	60
25	35	38	34	40	56
10	30	33	30	35	53
WRITING					
90	66	70	67	74	85
75	57	61	57	66	82
50	46	51	46	53	78
25	36	41	37	42	73
10	29	34	30	35	68

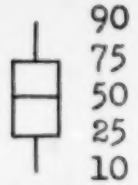
Table 15

Converted Scores Corresponding to Selected Percentile Ranks: SPANISH SUBGROUPS

LISTENING COMPREHENSION	CHILE	COLOMBIA	SPAIN
Percentile Ranks	Converted Scores		
90	59	59	59
75	55	58	57
50	52	55	54
25	48	53	52
10	47	49	46
SPEAKING			
90	97	106	109
75	92	102	104
50	86	97	99
25	82	92	93
10	75	86	90
READING			
90	64	66	66
75	62	65	64
50	59	61	61
25	55	57	56
10	52	54	51
WRITING			
90	82	86	86
75	79	82	81
50	76	80	77
25	71	76	72
10	68	71	68

Figure 1. MLA Listening Comprehension Tests

Converted scores for selected percentile ranks:



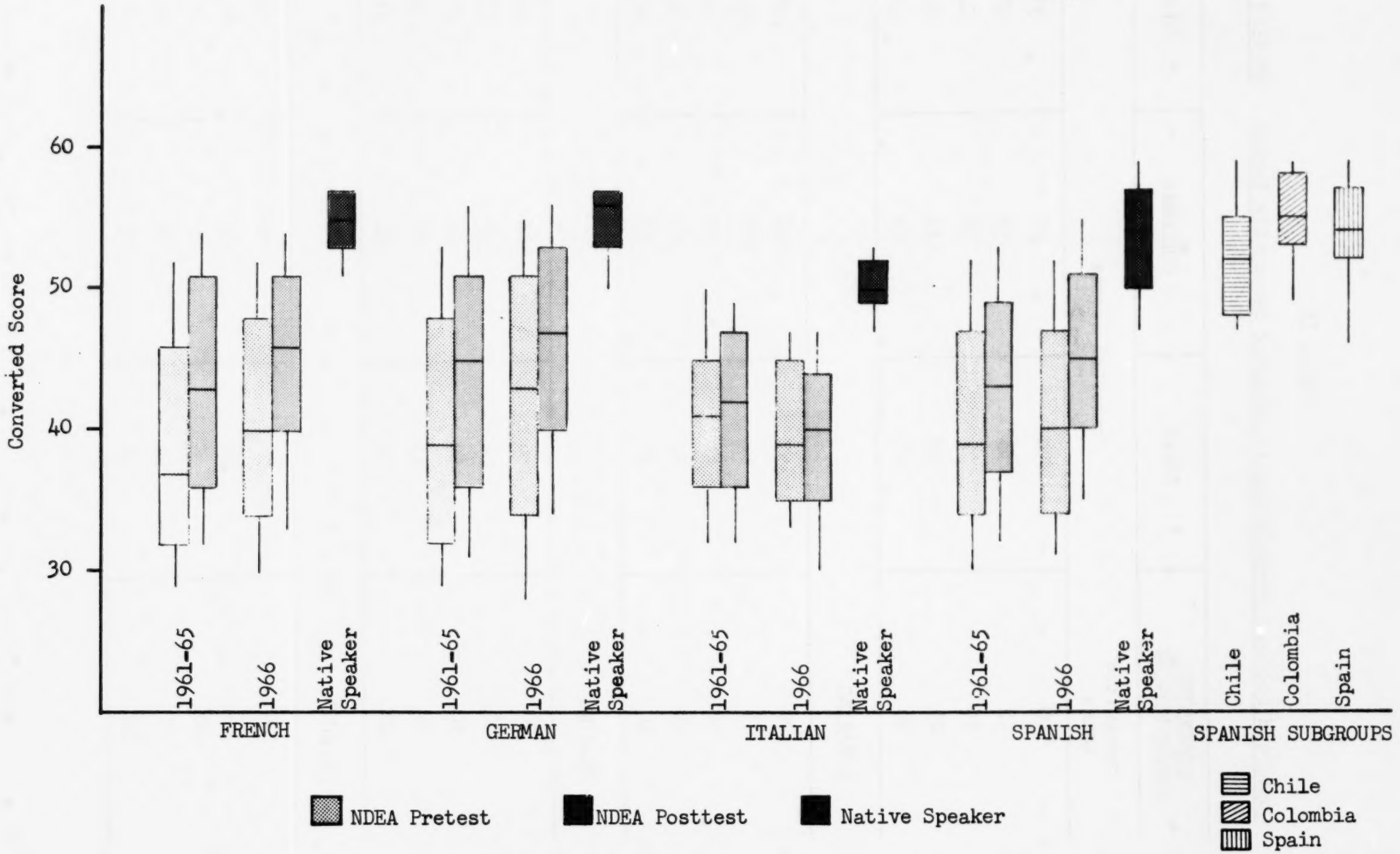
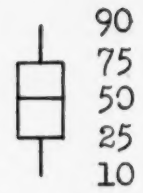


Figure 2. MLA Speaking Tests

Converted scores for selected percentile ranks:



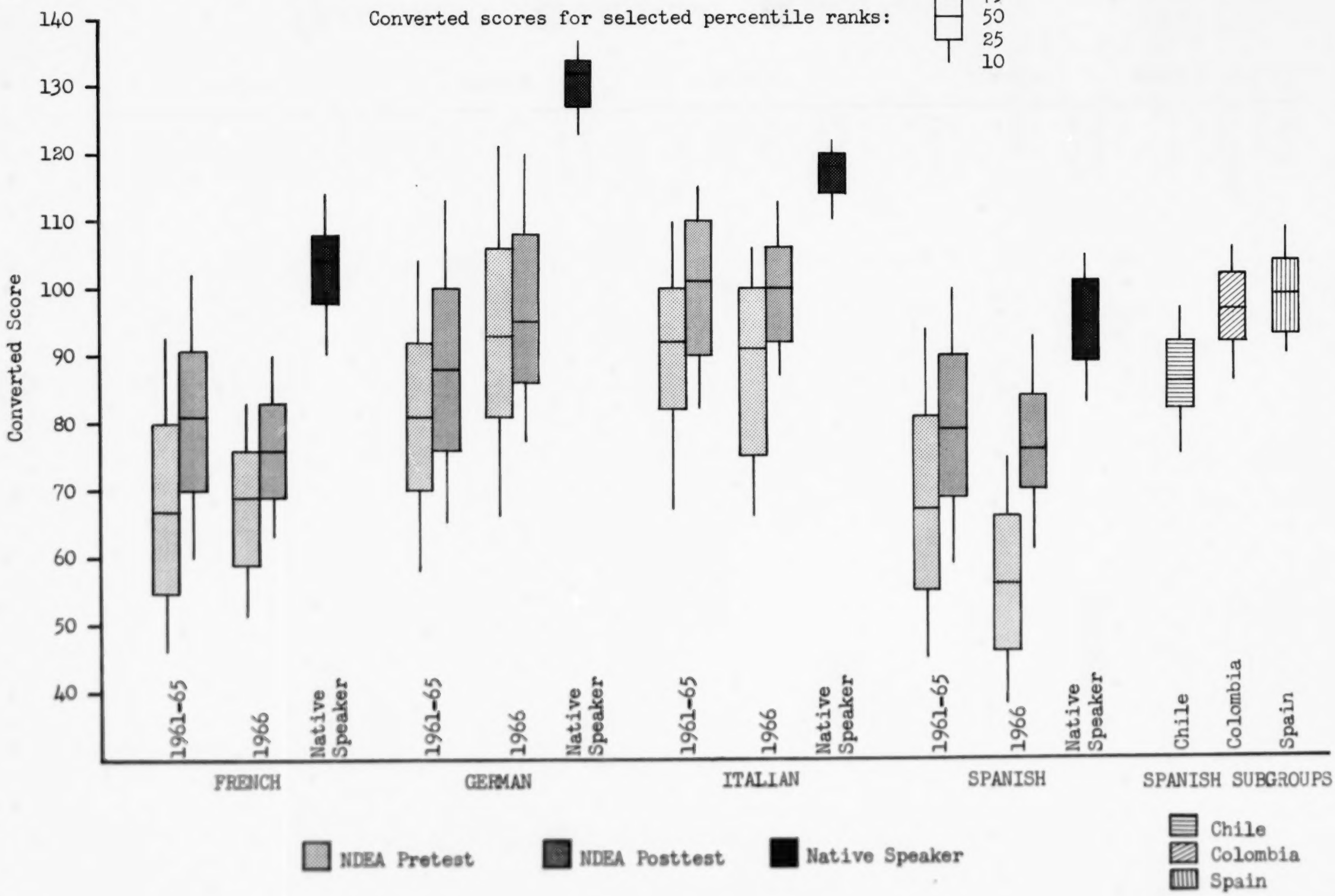


Figure 3. MLA Reading Tests

Converted scores for selected percentile ranks: 90 75 50 25 10

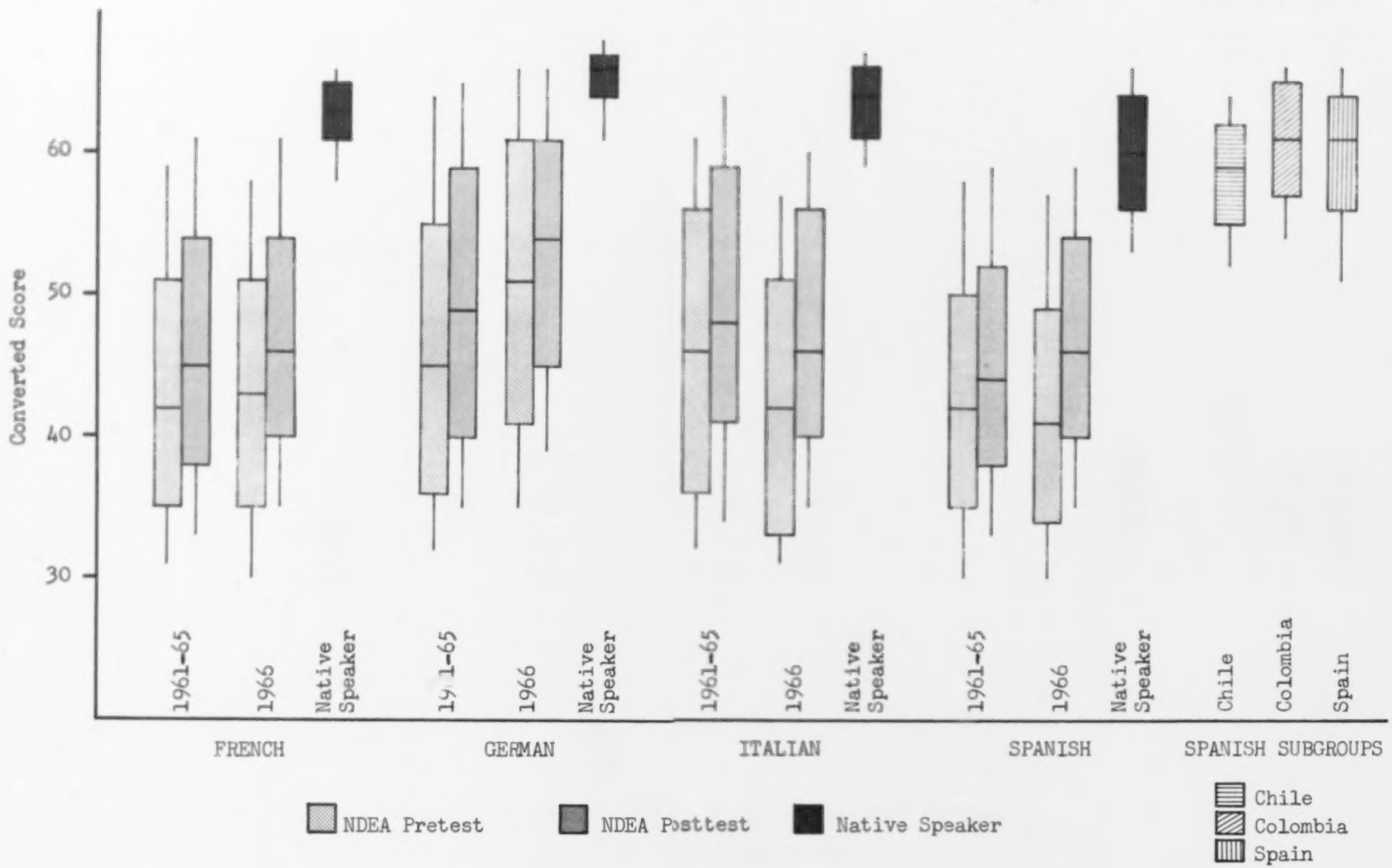
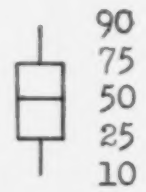


Figure 4. MLA Writing Tests

Converted scores for selected percentile ranks:



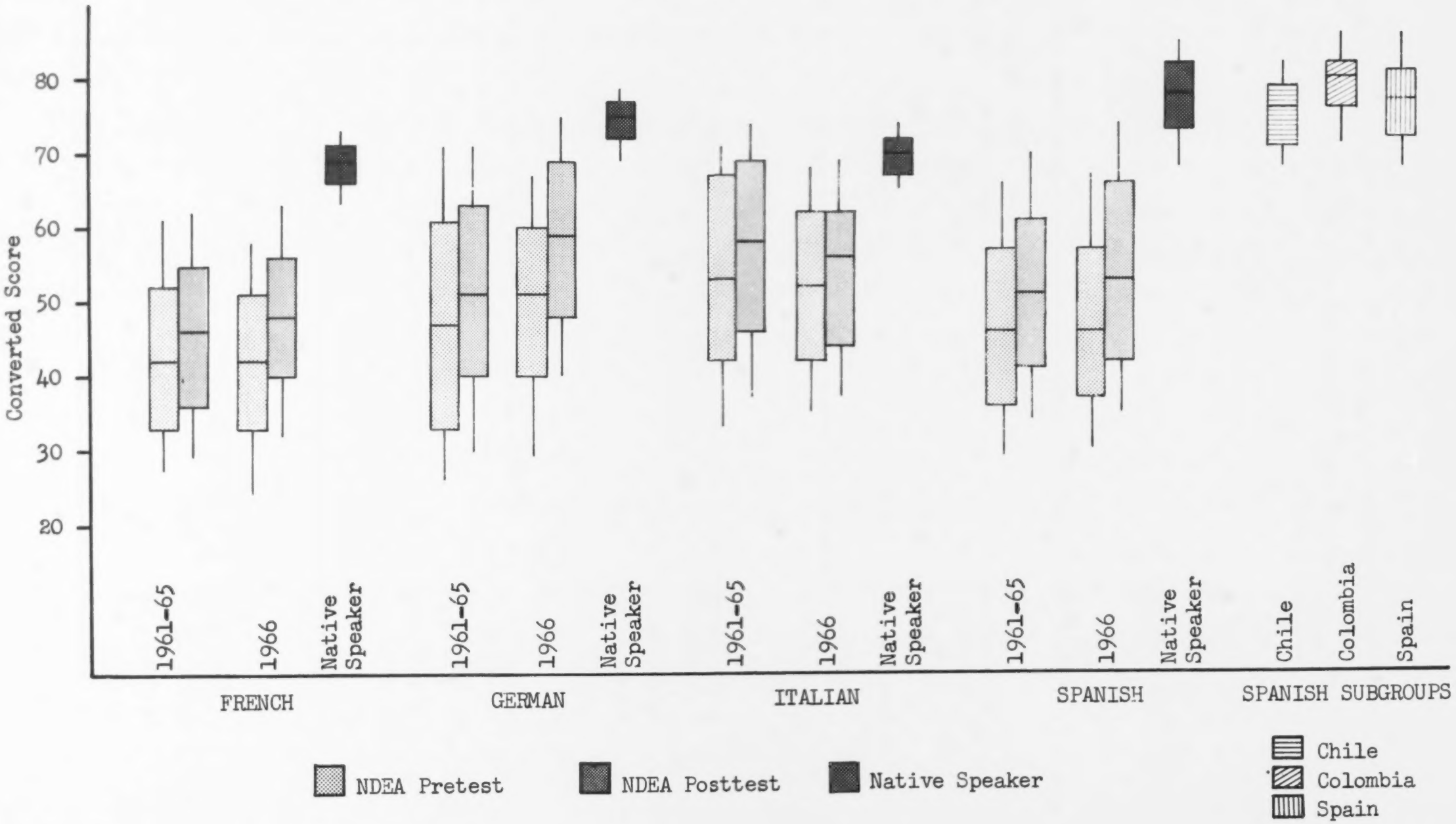


Table 16. Intercorrelation Tables

Test	Listening	Speaking	Reading	Writing	Mean	St. Dev.
Sample: Native Speaker FRENCH (N = 289)						
Listening		.128	.361	.369	54.5	2.8
Speaking	.128		.084	.160	102.8	8.5
Reading	.361	.084		.424	62.6	3.1
Writing	.369	.160	.424		68.3	4.1
Sample: FRENCH, NDEA Posttest 1961 (Test Analysis Sample)						
Listening		.800	.800	.797	42.3	8.7
Speaking	.800		.736	.782	84.1	18.6
Reading	.800	.736		.858	46.7	10.5
Writing	.797	.782	.858		46.5	12.5
Sample: Native Speaker GERMAN (N = 298)						
Listening		.137	.411	.351	54.5	2.8
Speaking	.137		.126	.215	130.4	5.4
Reading	.411	.126		.415	65.2	3.2
Writing	.351	.215	.415		74.2	3.9
Sample: Native Speaker ITALIAN (N = 226)						
Listening		.032	.260	.247	50.3	2.1
Speaking	.032		.039	.079	116.7	5.1
Reading	.260	.039		.434	63.5	3.3
Writing	.247	.079	.434		70.0	3.5
Sample: Native Speaker SPANISH (N = 270)						
Listening		.189	.344	.316	52.4	6.5
Speaking	.189		.162	.200	94.1	9.5
Reading	.344	.162		.474	59.6	5.4
Writing	.316	.200	.474		77.0	7.0

Table 17. Item Analysis Specifications and Identification

IA Series	Form	Test Title	Items	Test Code	Criterion (R) Items	Criterion (R) Type	Dropout
155-1	OML1	French Reading	1-50	NS-F-RD	1-50	IS50	Yes
155-2	OML1	French Writing	1-60	NS-F-WR	1-60	IS60	Yes
155-3	OML1	French Listening	1-36	NS-F-LC	1-36	IS36	---
155-4	OML1	French Speaking	1-53	NS-F-SPK	1-40	IS40	---
156-1	OML2	German Reading	1-50	NS-G-RD	1-50	IS50	Yes
156-2	OML2	German Writing	1-60	NS-G-WR	1-60	IS60	Yes
156-3	OML2	German Listening	1-36	NS-G-LC	1-36	IS36	---
156-4	OML2	German Speaking	1-53	NS-G-SPK	1-40	IS40	---
157-1	OML1	Italian Reading	1-50	NS-I-RD	1-50	IS50	Yes
157-2	OML1	Italian Writing	1-60	NS-I-WR	1-60	IS60	Yes
157-3	OML1	Italian Listening	1-36	NS-I-LC	1-36	IS36	---
157-4	OML1	Italian Speaking	1-53	NS-I-SPK	1-40	IS40	---
161-1	OML2	Spanish Reading	1-50	NS-S-RD	1-50	IS50	Yes
161-2	OML2	Spanish Writing	1-60	NS-S-WR	1-60	IS60	Yes
161-3	OML2	Spanish Listening	1-36	NS-S-LC	1-36	IS36	---
161-4	OML2	Spanish Speaking	1-53	NS-S-SPK	1-40	IS40	---
		Sample for Chile					
158-1	OML2	Spanish Reading	1-50	CH-S-RD	1-50	IS50	Yes
158-2	OML2	Spanish Writing	1-60	CH-S-WR	1-60	IS60	Yes
158-3	OML2	Spanish Listening	1-36	CH-S-LC	1-36	IS36	---
158-4	OML2	Spanish Speaking	1-53	CH-S-SPK	1-40	IS40	---
		Sample for Colombia					
159-1	OML2	Spanish Reading	1-50	CO-S-RD	1-50	IS50	Yes
159-2	OML2	Spanish Writing	1-60	CO-S-WR	1-60	IS60	Yes
159-3	OML2	Spanish Listening	1-36	CO-S-LC	1-36	IS36	---
159-4	OML2	Spanish Speaking	1-53	CO-S-SPK	1-40	IS40	---
		Sample for Spain					
160-1	OML2	Spanish Reading	1-50	SP-S-RD	1-50	IS50	Yes
160-2	OML2	Spanish Writing	1-60	SP-S-WR	1-60	IS60	Yes
160-3	OML2	Spanish Listening	1-36	SP-S-LC	1-36	IS36	---
160-4	OML2	Spanish Speaking	1-53	SP-S-SPK	1-40	IS40	---

Table 18

Frequency Distributions of Item Statistics: FRENCH (Form OML1)

Comparison of Results for Test Analysis Sample (TA) and Native Speaker Sample (NS)

Delta	LISTENING		SPEAKING*		READING		WRITING	
	TA	NS	TA	NS	TA	NS	TA	NS
Over 17							9	
16.0-16.9	1				2		3	1
15.0-15.9	1		2		2		4	1
14.0-14.9	3		3		9	1	7	
13.0-13.9	13		3	1	6	2	9	3
12.0-12.9	9	1	3	1	11	2	12	1
11.0-11.9	4		5	7	12	2	4	5
10.0-10.9	2	3	12	6	7	4	7	3
9.0- 9.9	3	2	6	9	1	5	3	6
8.0- 8.9		2	5	4		1	2	5
7.0- 7.9		5		5		6		3
6.0- 6.9		7		1		2		7
Below 6		16	1	6		25		25
Number of Items	36	36	40	40	50	50	60	60
Mean Δ	12.7	7.2	11.0	9.0	12.6	7.8	13.4	8.1
S.D. Δ	1.6	1.7		2.2	1.7	2.5	2.6	2.8
r-biserial								
Over .70	9		10		6		15	
.60-.69	10		18	1	14	2	22	
.50-.59	13	4	7	2	14	6	15	5
.40-.49	2	12	1	13	12	4	4	14
.30-.39	1	4	1	9	3	6	3	14
.20-.29	1		2	5	1	5	1	2
.10-.19				3		2		
Below .10				1				
Number of Items	36	20	39	34	50	25	60	35
Not computed		16		6		25		25
Mean r-bis	.61	.45	.62	.36	.56	.40	.62	.40
S.D. r-bis	.13	.06		.12	.12	.14	.13	.08

*The criterion score for the Native Speaker (NS) item analysis was the score on items 1-40.

Table 19

Frequency Distributions of Item Statistics: GERMAN (Form OML2)

Comparison of Results for Test Analysis Sample (TA) and Native Speaker Sample (NS)

Delta	LISTENING		SPEAKING*		READING		WRITING	
	TA	NS	TA	NS	TA	NS	RA	NS
Over 17							3	1
16.0-16.9			1				3	
15.0-15.9			4		2		6	
14.0-14.9			1		2		9	2
13.0-13.9	2		4		8	1	8	
12.0-12.9	6	2	7	1	9	1	11	5
11.0-11.9	6		2	1	9	2	6	1
10.0-10.9	5		3	1	5	2	7	4
9.0- 9.9	5	2	4		5	3	4	4
8.0- 8.9	6	4	1	3	4	6	1	8
7.0- 7.9	2	3	3	4	5	3	1	12
6.0- 6.9	3	9	2	3	1	5	1	2
Below 6	1	16	8	27		27		21
Number of Items	36	36	40	40	50	50	60	60
Mean Δ	10.0	7.0	10.5	6.1	11.2	7.3	12.9	8.2
S.D. Δ	2.1	1.7		1.9	2.2	2.0	2.5	2.6
r-biserial								
Over .70	3		1		11		23	
.60-.69	11		4	4	14	1	24	1
.50-.59	13	4	4	2	10	5	11	2
.40-.49	4	9	10	4	8	5	1	16
.30-.39	1	5	6	2	5	10		10
.20-.29	3	2	5	1	1	1	1	5
.10-.19			1		1			5
Below .10			1			1		
Number of Items	35	20	32	13	50	23	60	39
Not computed	1	16	8	27		27		21
Mean r-bis	.55	.42	.42	.49	.57	.41	.66	.37
S.D. r-bis	.13	.08		.11	.15	.11	.11	.12

*The criterion score for the Native Speaker (NS) item analysis was the score on items 1-40.

Table 20

Frequency Distributions of Item Statistics: ITALIAN (Form OML1)

Comparison of Results for Test Analysis Sample (TA) and Native Speaker Sample (NS)

Delta	LISTENING		SPEAKING**		READING		WRITING	
	TA*	NS	TA	NS	TA	NS	TA	NS
Over 17		2						2
16.0-16.9								1
15.0-15.9		1						1
14.0-14.9								1
13.0-13.9				1				1
12.0-12.9		2		2		6		4
11.0-11.9				2		3		4
10.0-10.9		1		3		2		2
9.0- 9.9		3		4		3		1
8.0- 8.9		4		2		6		5
7.0- 7.9		8		1		8		7
6.0- 6.9		5		2		5		8
Below 6		10		23		17		23
Number of Items	36		40		50		60	
Mean Δ	8.3		7.1		8.0		8.4	
S.D. Δ	3.1		2.8		2.3		3.2	
r-biserial								
Over .70				3				
.60-.69				4				
.50-.59		1		4		7		3
.40-.49		9		4		6		11
.30-.39		12		2		12		11
.20-.29		4				6		11
.10-.19						2		1
Below .10								
Number of Items	26		17		33		37	
Not computed	10		23		17		23	
Mean r-bis	.37		.56		.38		.35	
S.D. r-bis	.08		.11		.11		.10	

*The Italian tests were not analyzed because there were too few cases.

**The criterion score for the Native Speaker (NS) item analysis was the score on items 1-40.

Table 21

Frequency Distributions of Item Statistics: SPANISH (Total Group)

Comparison of Results for Test Analysis Sample (TA) and Native Speaker Sample (NS)

Delta	LISTENING		SPEAKING*		READING		WRITING	
	TA	NS	TA	NS	TA	NS	TA	NS
Over 17		1			2		6	
16.0-16.9	4				3		11	
15.0-15.9	1		1		2		6	
14.0-14.9	6	1	3		6		13	3
13.0-13.9	8	2	3		13	4	9	1
12.0-12.9	6	2	8		11	3	8	3
11.0-11.9	6	3	3	2	7	2	3	10
10.0-10.9	3	4	2	7	4	6	1	6
9.0- 9.9	2	8	5	9	1	10	2	10
8.0- 8.9		7	6	8	1	3	1	8
7.0- 7.9		4	6	11		9		8
6.0- 6.9		3	1	1		2		2
Below 6		1	2	2		11		9
Number of Items	36	36	40	40	50	50	60	60
Mean Δ	13.1	9.7	10.4	8.7	13.1	8.9	14.5	9.3
S.D. Δ	1.9	2.4		1.5	2.0	2.4	2.3	2.4
r-biserial								
Over .70	3	3	5		3	1	22	
.60-.69	12	2	7	6	13	2	20	1
.50-.59	7	9	11	10	11	11	8	8
.40-.49	6	12	13	11	15	9	4	14
.30-.39	4	5	1	7	8	10	5	16
.20-.29	2	3	1	4		4	1	8
.10-.19	1					1		4
Below .10	1	1				1		
Number of Items	36	35	38	38	50	39	60	51
Not computed		1	2	2		11		9
Mean r-bis	.52	.47	.54	.46	.53	.43	.62	.38
S.D. r-bis	.17	.15		.12	.12	.13	.14	.12

*The criterion score for the Native Speaker (NS) item analysis was the score on items 1-40.

Table 22

Frequency Distributions of Item Statistics: SPANISH Subgroups*

Delta	LISTENING			SPEAKING (1-40)**			READING			WRITING		
	CH	CO	SP	CH	CO	SP	CH	CO	SP	CH	CO	SP
Over 17	1	1	1									
16.0-16.9							1			1		
15.0-15.9			1				1			2	1	1
14.0-14.9	1		1	1			1		1	1	1	2
13.0-13.9	5	3	1	1			3	1	3	3		2
12.0-12.9	3	2	2	4	3		5	6	3	4	6	6
11.0-11.9	2	1	3	6	1	1	1	3	4	5	5	5
10.0-10.9	6	3	6	8	6	1	3	2	7	11	4	9
9.0- 9.9	8	6	8	8	9	1	9	6	4	11	13	7
8.0- 8.9	3	6	2	4	9	10	6	11	5	5	8	7
7.0- 7.9	2	8	6	7	4	8	5	6	9	6	6	9
6.0- 6.9	4	1	4		4	7	5	3	3	3	3	4
Below 6	1	5	1	1	4	12	10	12	11	8	13	8
Number of Items	36	36	36	40	40	40	50	50	50	60	60	60
Mean Δ	10.3	9.0	9.8	9.8	8.7	7.0	9.2	8.5	8.8	9.7	8.9	9.4
S.D. Δ	2.6	2.5	2.6	1.9	1.9	1.6	2.8	2.3	2.5	2.6	2.4	2.5
r-biserial												
Over .70			3	1			2	4	4		2	1
.60-.69		3	5	1	6	2	2		12	3	3	2
.50-.69	4	3	6	10	5	5	11	11	4	4	11	13
.40-.49	7	13	11	8	8	5	7	7	4	15	7	14
.30-.39	13	3	4	11	7	9	6	9	10	11	11	14
.20-.29	8	6	4	3	6	3	7	4	2	12	10	3
.10-.19	1	2	1	2	2	3	3	2	2	4	2	3
Below .10	2	1	1	3	2	1	2	1	1	3	1	2
Number of Items	35	31	35	39	36	28	40	38	39	52	47	52
Not computed	1	5	1	1	4	12	10	12	11	8	13	8
Mean r-bis	.33	.38	.47	.40	.40	.38	.41	.44	.49	.34	.41	.41
S.D. r-bis	.14	.14	.17	.16	.16	.15	.19	.19	.18	.16	.16	.15

* CH = Chile
CO = Colombia
SP = Spain

**The criterion score for the Native Speaker (NS) item analysis was the score on items 1-40.

Table 23. Summary of Ratings on the FRENCH, GERMAN, and ITALIAN Speaking Tests

Picture-Item	FRENCH (N = 285) 0	1	2	3	4	5	Mean	GERMAN (N = 295) 0	1	2	3	4	5	Mean	ITALIAN (N = 225) 0	1	2	3	4	5	Mean
41 (Global)	1			60	207	17	3.84				1	196	98	4.33					99	126	4.56
Vocabulary																					
I - 42			12	100	132	41	3.71				3	109	183	4.61		3		1	20	201	4.85
II - 46			11	83	149	42	3.78					90	205	4.69					9	216	4.96
III - 50		6	32	142	82	23	3.29				5	100	190	4.63					18	207	4.92
Total V		6	55	325	363	106	3.59				8	299	578	4.64		3		1	47	624	4.91
Pronunciation																					
I - 43			1	53	177	54	4.00	1				86	208	4.69		3		1	91	130	4.53
II - 47			2	52	183	48	3.97	1			1	143	150	4.49				1	102	122	4.54
III - 51			2	55	188	40	3.93				2	141	152	4.51					99	126	4.56
Total P			5	160	548	142	3.97	2			3	370	510	4.56		3		2	292	378	4.54
Structure																					
I - 44	1		5	85	153	41	3.80	2			1	98	194	4.63		3		1	24	197	4.83
II - 48				77	164	44	3.88				2	113	180	4.60					16	209	4.93
III - 52			6	107	145	27	3.68				1	100	194	4.65					30	195	4.87
Total S	1		11	269	462	112	3.79	2			4	311	568	4.63		3		1	70	601	4.88
Fluency																					
I - 45			11	110	137	27	3.63	2			6	136	151	4.46		3		3	101	118	4.47
II - 49			8	88	141	48	3.80				2	106	187	4.63					62	163	4.72
III - 53		3	48	126	94	14	3.24				18	114	163	4.49					5	120	100	4.42
Total F		3	67	324	372	89	3.56	2			26	356	501	4.53		3			8	283	381	4.54

*Ratings of zero are invalid. These zero ratings indicate failure of the professional scorer to grid the rating on the answer sheet. Their effect on the mean ratings is to reduce them less than 0.03.

Table 24. Comparison of the Ratings for the Three Spanish Groups on SPANISH SPEAKING

Picture-Item	CHILE (N = 75)							COLOMBIA (N = 90)							SPAIN (N = 100)						
	0	1	2	3	4	5	Mean	0	1	2	3	4	5	Mean	0	1	2	3	4	5	Mean
41 (Global)	1		10	35	25	4	3.27				20	68	2	3.80				9	70	21	4.12
Vocabulary																					
I - 42		3	17	39	14	2	2.93		2	11	35	38	4	3.34		3	21	37	36	3	3.15
II - 46		2	14	49	9	1	2.91		1	6	41	34	8	3.47			14	49	31	6	3.29
III - 50		3	22	30	19	1	2.91		1	7	36	43	3	3.44		5	33	32	23	7	2.94
Total V		8	53	118	42	4	2.92		4	24	112	115	15	3.42		8	68	118	90	16	3.13
Pronunciation																					
I - 43			4	32	37	2	3.49			1	12	45	32	4.20		1			44	55	4.52
II - 47		1	5	32	30	7	3.49		1		9	40	40	4.31					37	63	4.63
III - 51		1	4	40	25	5	3.39		2	2	9	43	34	4.17				2	40	58	4.56
Total P		2	13	104	92	14	3.46		3	3	30	128	106	4.23		1		2	121	176	4.57
Structure																					
I - 44			3	36	35	1	3.45			1	18	54	17	3.97		1	3	24	50	22	3.89
II - 48		1	4	33	35	2	3.44		1	1	8	60	20	4.08			4	9	69	18	4.01
III - 52		1	6	34	33	1	3.36		2	2	22	55	9	3.74			8	29	46	17	3.72
Total S		2	13	103	103	4	3.42		3	4	48	169	46	3.93		1	15	62	165	57	3.87
Fluency																					
I - 45		2	10	26	32	5	3.37		1	4	23	30	32	3.98		1	11	35	34	19	3.59
II - 49		2	5	25	40	3	3.49		1	6	17	39	27	3.94			5	26	49	20	3.84
III - 53		4	9	28	29	5	3.29		2	4	18	37	29	3.97		1	18	38	28	15	3.38
Total F		8	24	79	101	13	3.39		4	14	58	106	88	3.96		2	34	99	111	54	3.60